

Claims

1. Method for acquiring data from machine-readable documents, the data being assigned to a database, by extracting individual data from the document as automatically as possible and entering them into corresponding database fields, and, if data cannot be extracted from a document with the required degree of reliability for one or more particular database fields, executing the following steps:

- displaying of the document on a display screen,
- indication on the display screen of the database field for which the data cannot be extracted with the necessary degree of reliability,
- execution of a proposal routine with which string sections in the vicinity of a pointer on the display screen that can be moved by a user are selected, marked, and proposed for extraction.

2. Method according to Claim 1,

characterized in that the string section is selected, marked, and proposed for extraction in accordance with concept information assigned to the database field.

3. Method according to Claim 2,

characterized in that the concept information describes the syntax and/or the semantics of the database field, so that the proposal routine selects and marks a string section that is to be marked in a manner corresponding to the syntax or to the semantics of the respective database field.

4. Method according to Claim 3,

characterized in that the information concerning syntax describes the number of numerals and/or letters and/or predetermined formats of the string section that is to be read.

5. Method according to Claim 3 or 4,

characterized in that

the information concerning semantics describes specified terms, for example using a lexicon.

6. Method according to one of Claims 1 to 5,

characterized in that a string section is selected that is situated between two limiting [or: boundary] characters.

7. Method according to Claim 6,

characterized in that the limiting characters include empty characters and/or punctuation marks.

8. Method according to one of Claims 1 to 7,

characterized in that the text of documents in graphic representation is first converted into coded text using an OCR method, and the proposal routine represents, in addition to the marked string section in graphic representation, the coded text of this string section.

9. Method according to one of Claims 1 to 7,

characterized in that in addition to the marked string section, this string section is displayed again on the display screen in an enlarged representation.

10. Method according to one of Claims 1 to 9,

characterized in that after the marking of a string section, the proposal routine activates a function with which the content of the marked string section is transferred into the database through the actuation of one or more predetermined keys.

11. Method according to one of Claims 1 to 10,

characterized in that during the execution of the proposal routine, after the movement of the pointer a predetermined time wait interval is observed, during which the pointer must not be moved, before a string section is selected.

12. Method for acquiring data from machine-readable documents, the data being assigned to a database, in particular according to one of Claims 1-11,

characterized in that after data have been read from a first table row into corresponding database fields, the further table entries are automatically determined through a comparison of string sections situated under the last table row with the string sections of the first table

row, which correspond to the read-in data, and these additional table entries are automatically extracted.

13. Method according to Claim 12,
characterized in that the comparison between the string sections takes place using a string matching method.

14. Method according to Claim 12 or 13,
characterized in that the determined table entries are marked.

15. Method according to Claim 14,
characterized in that functions are provided for editing the marked table entries.

16. System for acquiring data from machine-readable documents, comprising a computer (12) having a storage device (13) and a CPU (14), a software product for executing the method according to one of Claims 1-15 being stored in the storage device (13).

17. System according to Claim 16,
characterized in that the system has an input device (17) in the form of a mouse (6) and/or keyboard (7).

18. System as recited in Claim 16 or 17,
characterized in that the system has a scanner (16) for the optical scanning of documents.

19. Computer program products that, in its storing and execution on a computer (12), effects a method according to one of Claims 1 to 15.